# Supplementary for VectorFloorSeg: Two-Stream Graph Attention Network for Vectorized Roughcast Floorplan Segmentation

Bingchen Yang[1], Haiyong Jiang[1§], Hao Pan[2], Jun Xiao[1*]
[1] School of Artificial Intelligence, University of Chinese Academy of Sciences
[2] Microsoft Research Asia
yangbingchen211@mails.ucas.ac.cn, haiyong.jiang@ucas.ac.cn
haopan@microsoft.com, xiaojun@ucas.ac.cn

In this document, we present extra details of graph construction, experimental analysis, and visual results.

## A. Algorithm Details

**Floorplan partition.** Floorplans generally have complex layouts containing complicated line sketches parsed from the input vector graphics. For example, as illustrated in the middle part of Fig. 1(I), two rectangles representing horizontal and vertical wall blocks cross each other; some of the wall segments are thus spatially overlapped, while some of the intersections between wall segments are not marked as endpoints. The overlapping segments and missing endpoints would pose challenges to boundary-based room segmentation. Besides, as is mentioned in Sec. 3.1., some rooms do not have consecutive wall segments as boundary, leading to ambiguous room partitions.

We simplify the problem by firstly employing split and merge on wall segments to remove the overlaps and make up for missing endpoints. A room area in the floorplan can thus be tightly represented by its boundary polygon.

Then we follow the intuition that a good room region is usually as regular and rectangular as possible, so that designers usually partition open rooms by extending existing wall structures. We thus extend wall segments that form the railings, wall corners and junctions. By viewing the floorplan line sketch as a graph (i.e. endpoints cast as graph vertices and wall segments be edges), we find these structures can be represented by specific types of graph vertices (i.e with degree two or three) and adjacent edges. As illustrated in middle and right part of Fig. 1(II), for a vertex in degree two, we simply extend its two adjacent edges until they intersect with other edges. For a vertex in degree three, we first calculate two-by-two angles between the three adjacent edges: if the largest angle of two edges is close to $\pi$, we extend the remaining edge until intersection; otherwise,

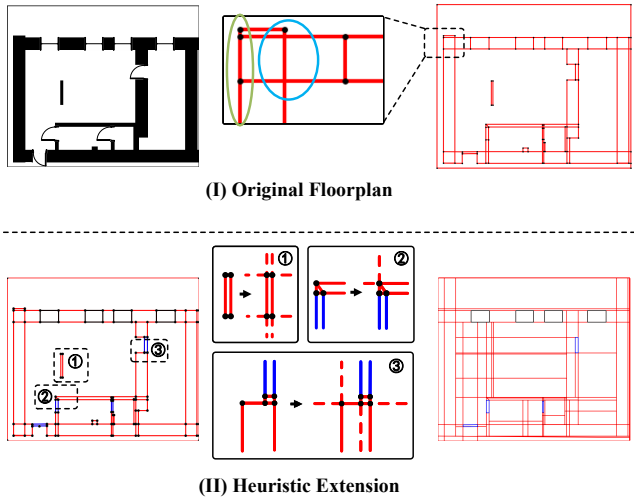


Figure 1. Details of floorplan processing and partitioning. (I)left: the input floorplan; middle: zoomed in to show the overlap and straddle of wall segments, circled by green and light blue ovals respectively; right: the line sketch parsed from input floorplan, where door curves and wall colors are discarded and only boundary segments are retained. (II)left: we make a split and merge on the floorplan line sketch to remove overlapping segments and recover missing endpoints; middle: zoomed in to show the extension mechanism in different circumstances, ①: on railings, ②: on corners, ③: on junctions; right: the extension result.

the two edges with largest angle are extended. Note that we do not extend the edges marked as window (in black stroke) or door (in blue), since these edges (or extensions) generally do not constitute room partition boundaries.

Finally after the line extensions, we make a split and merge pass again to simplify the whole graph.

**Floorplan rasterization and feature indexing.** For floorplan rasterization, we first extract the bounding box (bbox) of a vector floorplan and set the coordinates of the upper-left corner of the bbox to $(0, 0)$ and that of lower-right corner to

§Haiyong is the Project Lead.
*Corresponding Author.

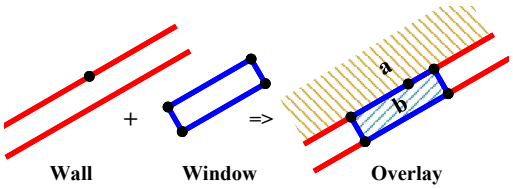


Figure 2. Non-standard drawing process leads to multiple primal edges in $\mathcal{E}_{v_i^*} \cap \mathcal{E}_{v_j^*}$.

$(h, w)$. The longer side of the bbox with length $\max(h, w)$ is mapped to $H - 1$, while the shorter side with length $\min(h, w)$ is mapped to $\frac{\min(h,w)}{\max(h,w)}(H-1)$, where we assume the rasterized image is of size $H \times H$. We keep the aspect ratio of the floorplan by padding zeros to margins.

We establish the relationship between the primal vertices $V$ (i.e., endpoints of line segments) and the rasterized floorplan image by vertex-to-pixel coordinate mapping, which calculates the relative position of a primal vertex w.r.t. the bbox. Without loss of generality we assume the longer side lies on the horizontal axis. For a vertex of coordinates $(x, y)$, the coordinate mapping function is defined as follows:

$$\begin{aligned} \text{map}(y) &= \left\lfloor \frac{y}{\max(h,w)}(H-1) + \frac{H-1}{2}\left(1 - \frac{\min(h,w)}{\max(h,w)}\right) \right\rfloor \\ \text{map}(x) &= \left\lfloor \frac{x}{\max(h,w)}(H-1) \right\rfloor \end{aligned} \tag{1}$$

where $\lfloor \cdot \rfloor$ denotes the floor operation, $\frac{H-1}{2}\left(1 - \frac{\min(h,w)}{\max(h,w)}\right)$ is the number of padded pixels on the vertical direction.

Given the rasterized floorplan, we feed it into a CNN backbone to learn image features, which is bilinearly interpolated into the same size as input. We get the feature vector for each primal vertex by indexing image features and adding a 2D positional encoding based on Eq. 1, where the positional encoding scheme follows [2] (see also Eq. 2 of main text).

**Multiple primal edges in $\mathcal{E}_{v_i^*} \cap \mathcal{E}_{v_j^*}$.** As there may be extra vertices on a wall due to non-standard drawing, some regions share more than one primal edge as shown in Fig. 2.

## B. More Experimental Results

In this part, we show more results on two evaluating datasets (i.e., R2V [5] and CubiCasa-5k [4]) using ResNet-101 [3] as the image backbone. We can see our method has consistent performance improvements comparing with other image-based segmentation methods. Note that our method enjoys a great advantage on RI due to direct segmentation on vector floorplans. Visualization of results from all comparing methods is presented in Fig. 3.

**Figure illustration for limitation.** As shown in Fig. 3 column 2,3, we mark the failure cases by dotted line boxes in three color, where green box denotes misclassification for extremely small regions, blue box denotes wrong label assignments for similar categories, and red box denotes reluctance for dealing with curves in the floorplan.

| Methods | Backbone | mIoU | mAcc | RI |
|---|---|---|---|---|
| DeepLabV3+ [1] | ResNet-101 | 79.66 | 88.18 | 76.62 |
| DNL [7] | – | 77.10 | 85.55 | 67.24 |
| UperNet [6] | – | 76.76 | 85.39 | 78.02 |
| OCRNet [8] | – | 77.98 | 85.34 | 70.82 |
| Ours | – | **81.38** | **89.86** | **86.20** |

Table 1. Comparison results on R2V.

| Methods | Backbone | *val*-set | | | *test*-set | | |
|---|---|---|---|---|---|---|---|
| | | mIoU | mAcc | RI | mIoU | mAcc | RI |
| DeepLabV3+ [1] | ResNet-101 | 62.11 | 74.49 | 52.43 | 59.45 | 72.78 | 51.43 |
| DNL [7] | – | 60.45 | 72.98 | 47.57 | 58.18 | 71.61 | 47.06 |
| UperNet [6] | – | 61.79 | 74.13 | 56.14 | 58.27 | 71.90 | 54.97 |
| OCRNet [8] | – | 60.44 | 72.94 | 43.99 | 57.13 | 70.62 | 41.89 |
| Ours | – | **64.36** | **76.98** | **69.55** | **62.49** | **75.48** | **67.51** |

Table 2. Comparison results on CubiCasa-5k.

## References


[1] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European conference on computer vision (ECCV)*, pages 801–818, 2018. 2

[2] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. OpenReview.net, 2021. 2

[3] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 2

[4] Ahti Kalervo, Juha Ylioinas, Markus Häikiö, Antti Karhu, and Juho Kannala. Cubicasa5k: A dataset and an improved multi-task model for floorplan image analysis. In *Scandinavian Conference on Image Analysis*, pages 28–40. Springer, 2019. 2, 4

[5] Chen Liu, Jiajun Wu, Pushmeet Kohli, and Yasutaka Furukawa. Raster-to-vector: Revisiting floorplan transformation. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2195–2203, 2017. 2, 4

[6] Tete Xiao, Yingcheng Liu, Bolei Zhou, Yuning Jiang, and Jian Sun. Unified perceptual parsing for scene understanding. In *Proceedings of the European conference on computer vision (ECCV)*, pages 418–434, 2018. 2

[7] Minghao Yin, Zhuliang Yao, Yue Cao, Xiu Li, Zheng Zhang, Stephen Lin, and Han Hu. Disentangled non-local neural networks. In *European Conference on Computer Vision*, pages 191–207. Springer, 2020. 2

[8] Yuhui Yuan, Xilin Chen, and Jingdong Wang. Object-contextual representations for semantic segmentation. In *European conference on computer vision*, pages 173–190. Springer, 2020. 2

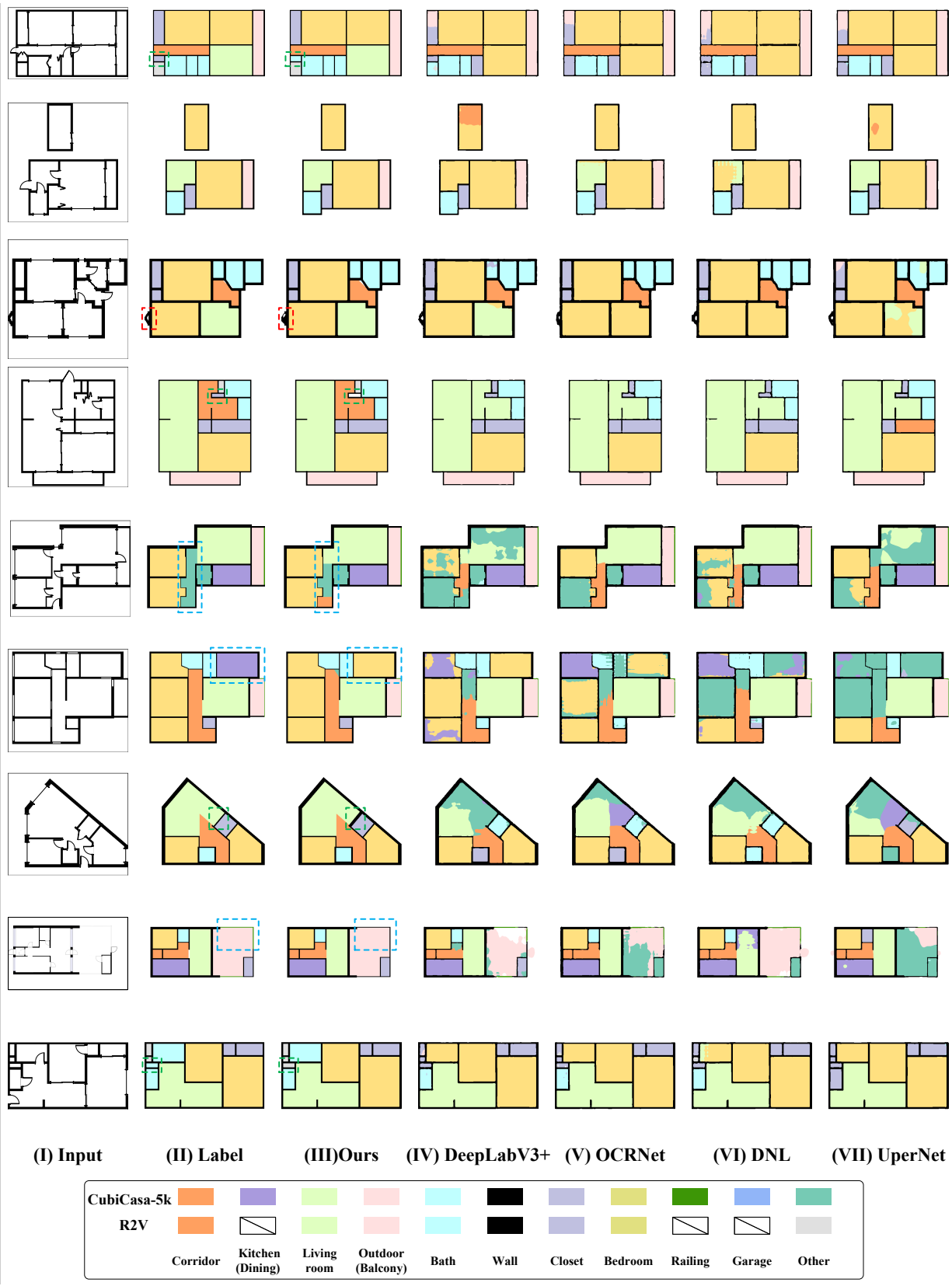


Figure 3. Qualitative results of all comparison methods on two floorplan datasets. Row 1-4, 9 denote results of R2V [5], row 5-8 denote that of CubiCasa-5k [4]. Best viewed by zooming in.